JOURNAL OF VIROLOGY, Oct. 2002, p. 10020–10023 Vol. 76, No. 19
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.19.10020–10023.2002

Lack of Canonical E6 and E7 Open Reading Frames in Bird Papillomaviruses: *Fringilla coelebs* Papillomavirus and *Psittacus erithacus timneh* Papillomavirus

Masanori Terai,[1] Rob DeSalle,[2] and Robert D. Burk[1,3,4]*

Departments of Microbiology and Immunology,[1] Pediatrics,[4] and Epidemiology and Social Medicine,[3] Comprehensive Cancer Center, Albert Einstein College of Medicine, Bronx, New York 10461, and Division of Invertebrate Zoology, American Museum of Natural History, New York, New York 10024[2]

Received 23 May 2002/Accepted 28 June 2002

Determination and analyses of the complete sequence of *Fringilla coelebs* papillomavirus and *Psittacus erithacus timneh* papillomavirus indicate that they represent a distinct and distant lineage of papillomaviruses. The lack of canonical E6-E7 open reading frames suggests that they serve adaptive functions during papillomavirus evolution.

Papillomaviruses (PVs) are a heterogeneous group of DNA viruses with closed-circular double-stranded DNA genomes about 8 kb in size. They are highly species-specific pathogens and cause benign and malignant lesions of squamous and mucosal epithelium in a wide range of animal species (22). The reported incidence of external neoplasms in wild birds is low, except for squamous papillomas found on the foot and lower leg of chaffinches (*Fringilla coelebs*) (13). The causative agent of these lesions was characterized as a PV based on the size and density of virus particles, physical properties of viral DNA, and analysis of capsid proteins by electrophoresis (20). The other known PV infection among avian species was in an African gray parrot (*Psittacus erithacus timneh*) (12). The PV genomes isolated from a chaffinch and an African gray parrot were cloned and partially sequenced (16, 18). We have determined and characterized the complete sequences of *F. coelebs* PV (FPV) and *P. erithacus timneh* PV (PePV).

The FPV genome was cloned into the *Eco*RI site of pBR328, and the PePV genome was cloned into the *Sal*I site of pBR322 (16, 18). To determine the nucleotide sequence, each cloned DNA was sequenced with the first primers selected from the vector sequence, and thereafter, additional primers were designed by sequence walking (7). Sequencing was performed in the Albert Einstein College of Medicine DNA sequencing core facility. The overlapping sequences were assembled manually. Several additional primers were designed and used to clarify sequence ambiguities. Once assembled, the sequence was analyzed for similarity to other PVs by using the basic local alignment sequence tool (BLAST) software (1). The same software was used to determine protein sequence similarities.

The assembled sequences revealed a total size of 7,729 bp for FPV and 7,304 bp for PePV. The complete sequences of the genomes are available from the GenBank database under accession no. AY057109 (FPV) and AF420235 (PePV). Examination of the FPV and PePV sequences for potential genes showed the typical complement of E1, E2, L2, and L1 open reading frames (ORFs) comparable in size with, and at positions similar to, those of other PVs, including overlaps between the E1 and E2 and the L2 and L1 ORFs. None of the small ORFs in FPV and PePV showed significant similarity with sequences for known E4 and E5 proteins. The PePV long control region contains one exact 12-bp E2-binding motif, whereas the FPV long control region contains four E2-binding motifs with one base pair mismatch. The predicted ORFs of FPV and PePV are shown in Fig. 1. The similarity of the FPV and PePV L1 ORF nucleotide and amino acid sequences was higher (nucleotide, 56.8%; amino acid, 55.8%) than the relationship of other ORFs, with similarity to E1 (56.1 and 44.0%, respectively) also elevated compared with that to E2 (48.6 and 36.3%, respectively) and L2 (45.6 and 27.8%, respectively). To investigate the relationship among FPV, PePV, and other PV genomes, the predicted amino acid sequences of the E1, E2, L2, and L1 ORFs were aligned with other representative PV sequences from each PV subfamily and unclassified genomes (8). The phylogenetic trees were created by using parsimony as implemented in the PAUP* program (version 4; D. L. Swofford, Sinauer Associates, Sunderland, Mass.) from published PV sequences available online from sexually transmitted disease sequence databases (http://hpv-web.lanl.gov/) and GenBank (4, 8, 10). We also used PV sequences, in part, from PVs that were sequenced in our laboratory, i.e., bovine PV type 3 (BPV3; GenBank accession no. AF486184), BPV5 (AF457465), equine PV (AF394740), and reindeer PV (AF443292) (15, 19, 21). Jackknife analysis was performed as in PAUP* with 1,000 jackknife replicates. Trees derived from the compiled amino acid sequences of the E1, E2, L1, and L2 proteins of the viral genome and from an analysis of the nucleotide sequences in the L1 gene are shown in Fig. 2. These trees indicated that there is significant homology between FPV and PePV but not with other classified or unclassified PV genomes. From the topology of the tree, it appears that FPV and PePV are the most distant PV sequences characterized to

* Corresponding author. Mailing address: Comprehensive Cancer Center, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461. Phone: (718) 430-3720. Fax: (718) 430-8975. E-mail: burk@aecom.yu.edu.

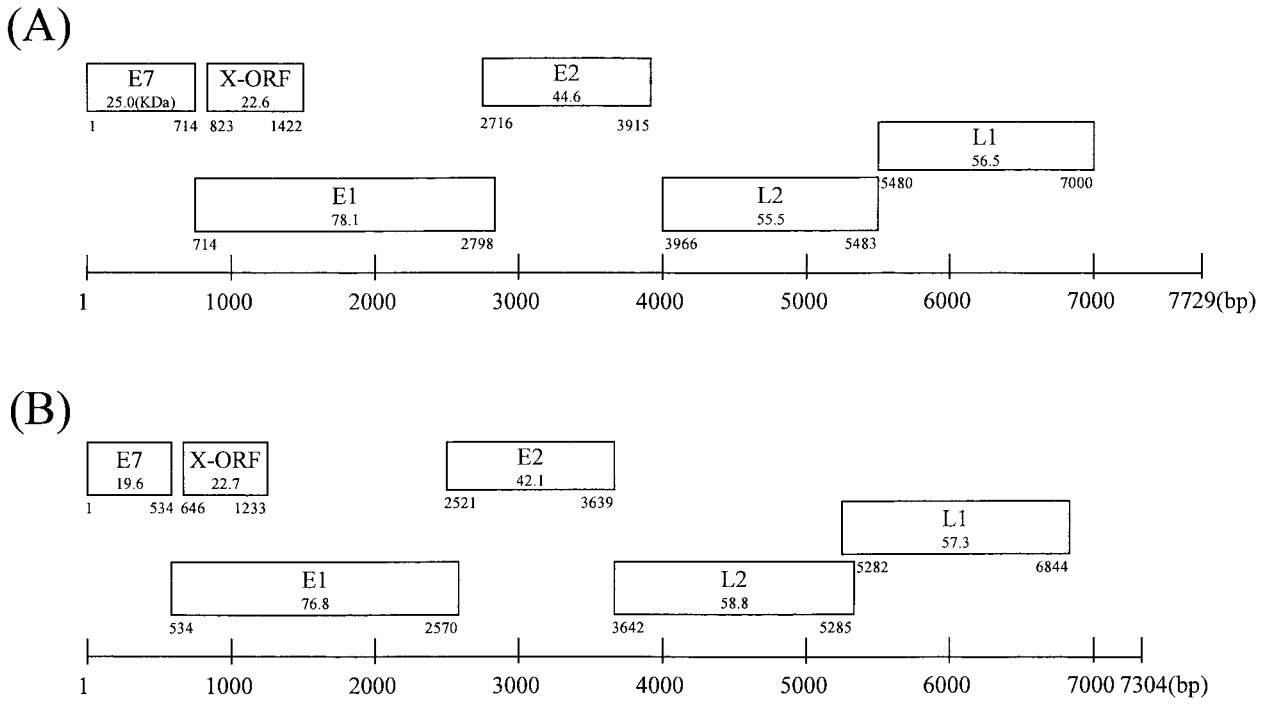

FIG. 1. Location of predicted ORFs of FPV (A) and PePV (B). Each ORF is represented as a rectangle. Numbers show nucleotide positions of the start and stop codons and predicted molecular masses of proteins (kilodaltons).

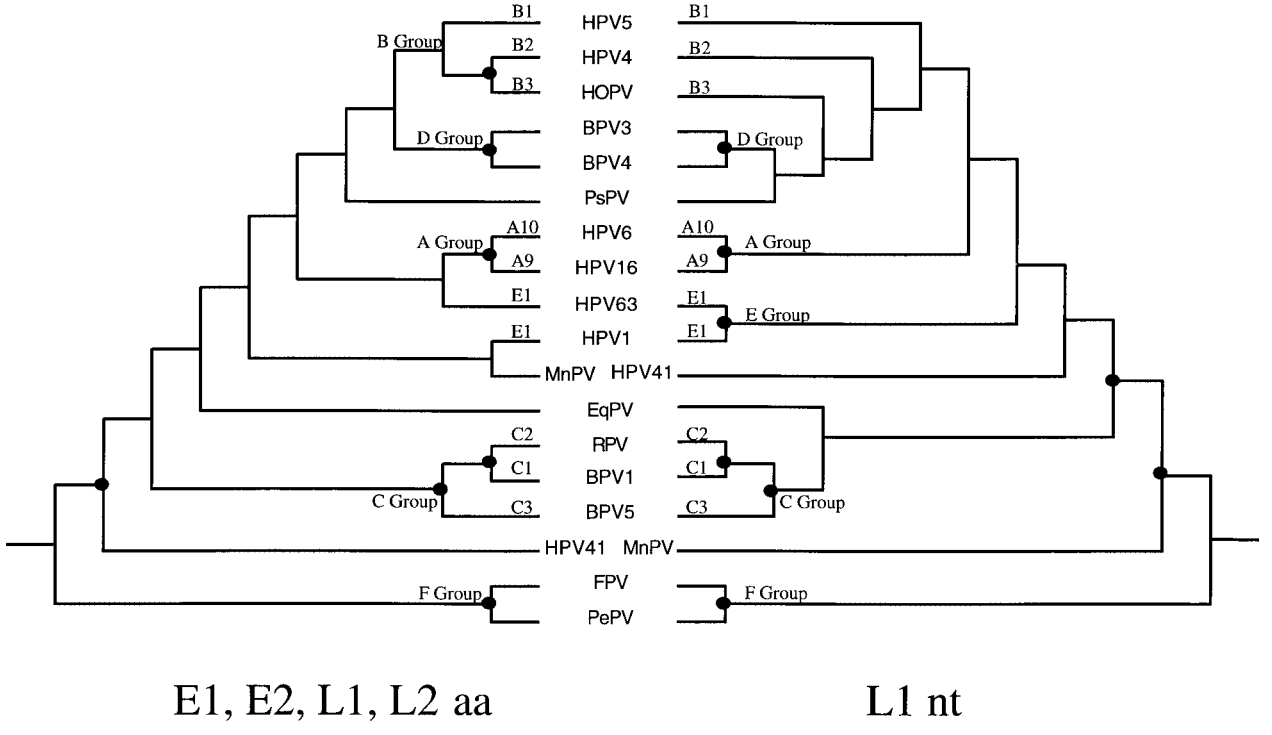

FIG. 2. Phylogenetic trees based on the alignment of the amino acid sequences (aa) of compiled ORFs (E1, E2, L1, and L2) of the indicated PV genomes (tree shown on the left) and based on aligned nucleotide sequences (nt) of just the L1 gene (shown on the right). The matrix used to generate the compiled ORF data tree had 2,862 aligned characters in it including gaps. Gaps were scored as missing in the parsimony analysis, and the resulting parsimony tree was 13,942 steps long with a confidence interval of 0.68 and a retention index (RI) of 0.38. The matrix used to generate the L1 nucleotide data tree had 2,088 aligned nucleotide characters in it including gaps. Gaps were scored as missing in the parsimony analysis, and the resulting parsimony tree was 7,401 steps long with a confidence interval of 0.39 and an RI of 0.32. Dots on the nodes in the trees indicate jackknife values for that node over 75%. The FPV and PePV sequences were compared with representative PV sequences (8). EqPV, equine PV; RPV, reindeer PV; HOPV, hamster oral PV; MnPV, *Mastomys natalensis* PV.

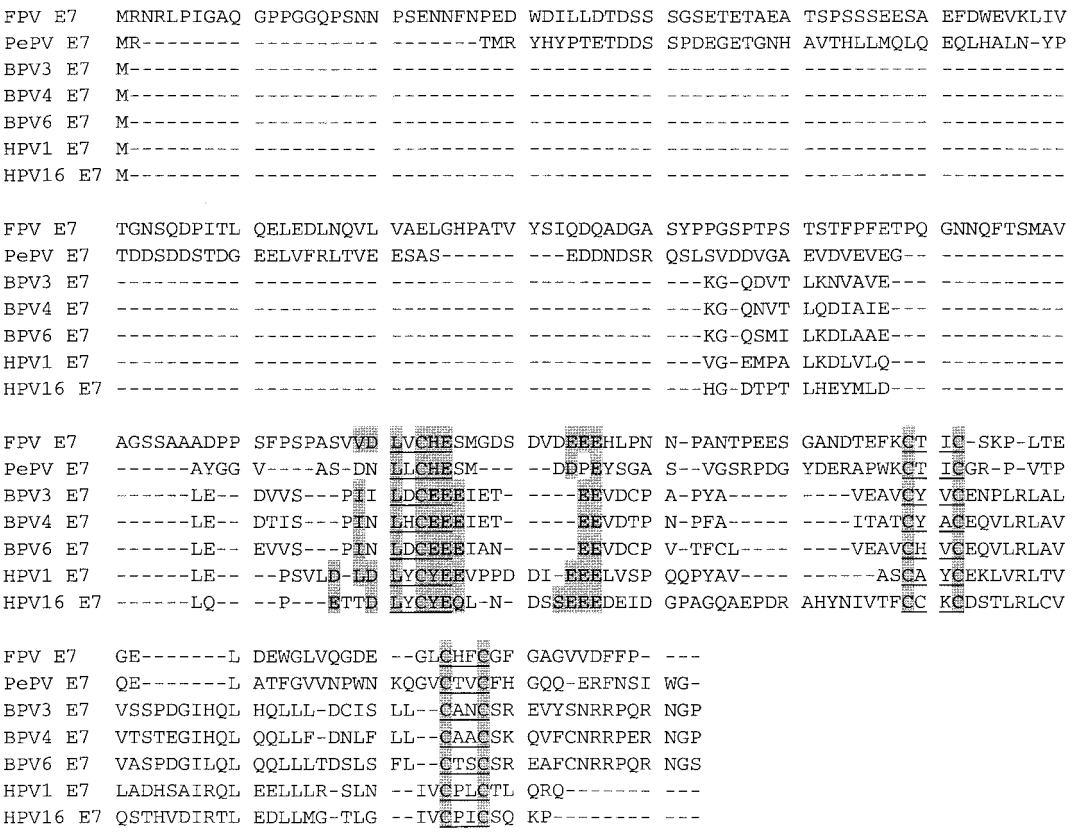

FIG. 3. Alignment of the putative pRB-binding motif in the amino acid sequences of the FPV, PePV, BPV3, BPV4, BPV6, HPV1, and HPV16 E7 ORFs. The consensus motif is shaded. A cysteine-X-X-cysteine (Cys-X-X-Cys; C-X-X-C) motif and a leucine-X-cysteine-X-glutamine (Leu-X-Cys-X-Glu; L-X-C-X-E) motif are underlined.

date and constitute a new subfamily, F. Close examination of the trees indicates a lack of robustness for many of the relationships inferred, but when a relationship is robust in one analysis, it is also robust in the other. These results indicate that classification of these viruses by using only the L1 nucleotide sequences or only amino acid sequences of the conserved regions of the viral genome may not be sufficient to resolve relationships robustly among the accepted subfamilies (4).

Although FPV and PePV lack canonical E6-E7 ORFs, two novel ORFs were identified proximal to the noncoding region and are labeled E7 and X-ORF. These ORFs had no significant similarity with sequences in existing databases but were similar to each other, i.e., the FPV and PePV E7 and X-ORF nucleotide sequences were 46.9 and 48.0% similar, respectively, whereas the amino acid sequences were 23.5 and 25.6% similar, respectively. The X-ORF ORF overlaps the E1 ORF and has no similarity to sequences for other known PV proteins. The E7 ORF occupies the region of the genome which normally contains the E6-E7 ORFs. Although the E7 ORF had no significant hits by BLAST, manual examination of the E7 ORFs indicated that they contain motifs conserved in the canonical E7 (24). The alignment of amino acid sequences of the FPV, PePV, human PV type 1 (HPV1), HPV16, BPV3, BPV4 and BPV6 E7 ORFs is shown in Fig. 3. The FPV and PePV E7 ORFs contain a leucine-X-cysteine-X-glutamine (Leu-X-Cys-X-Glu; L-X-C-X-E) motif reported to be critical for pRB binding and two cysteine-X-X-cysteine (Cys-X-X-Cys; C-X-X-C) motifs (2, 17, 24). Two such C-X-X-C motifs can stoichiometrically sequester a zinc ion through a tetrahedral arrangement of four cysteine sulfur ligands. Four copies of a C-X-X-C motif spaced at regular and invariant intervals are also contained in the canonical PV E6 ORF (3, 14). These conserved and repeated residues appear to be essential structures of the canonical E6 and can act as a multimerization domain (6). The FPV and PePV E7 ORFs might retain some vestigial functions of an ancestral E7 ORF, such as pRB binding. Based on the lack of significant similarity of the putative E7–X-ORF to the E6/E7, we suggest that the putative E7–X-ORF and E6-E7 ORFs have evolved adaptive functions for specialization within specific species from a common ancestral precursor (5). Interestingly, BPV3, BPV4, and BPV6 have E7 but not E6 ORFs which might have evolved, at least in part, through genomic rearrangements (9, 11). *Phocoena spinipinnis* PV (PsPV) causes genital warts in small cetaceans and has an E6 ORF containing four copies of the C-X-X-C motif spaced at regular and invariant intervals but lacks an identifiable E7 (23). We found sequences at positions 857 to 871 in PsPV that can be translated as Leu-Lys-Cys-Thr-Glu (L-K-C-T-E), which is part of the critical pRB binding domain, though the ORF was only 26 amino acids long and did not have an in-frame proximal ATG. Perhaps it is encoded from a spliced transcript. Taken together, FPV and PePV along with BPV3, BPV4, and BPV6

lack definable E6 ORFs, whereas PsPV and to a lesser extent FPV and PePV lack E7 ORFs, supporting the hypothesis that the E6 and E7 ORFs play a central function in adapting PV genomes to various species and tissues. In contrast, the E1, E2, L2, and L1 ORFs are well conserved in all PVs, and their products are concluded to be essential proteins for the PV life cycle. The characterization of the complete genomes of two avian species shows these to be the most divergent PV genomes analyzed to date.

Nucleotide sequence accession numbers. The newly assigned GenBank accession numbers for nucleotide and/or amino acid sequence data in this paper are AY057109 (FPV), AF420235 (PePV), AF486184 (BPV3), AF457465 (BPV5), AF394740 (equine PV), and AF443292 (reindeer PV).

This work was supported in part by grants from the NIH to R.D.B. We thank John Sundberg for providing reagents.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Barbosa, M. S., C. Edmonds, C. Fisher, J. T. Schiller, D. R. Lowy, and K. H. Vousden.** 1990. The region of the HPV E7 oncoprotein homologous to adenovirus E1a and SV40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. EMBO J. **9:**153–160.
3. **Barbosa, M. S., D. R. Lowy, and J. T. Schiller.** 1989. Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. J. Virol. **63:**1404–1407.
4. **Chan, S. Y., H. U. Bernard, M. Ratterree, T. A. Birkebak, A. J. Faras, and R. S. Ostrow.** 1997. Genomic diversity and evolution of papillomaviruses in rhesus monkeys. J. Virol. **71:**4938–4943.
5. **Cole, S. T., and O. Danos.** 1987. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. J. Mol. Biol. **193:**599–608.
6. **Danos, O., and M. Yaniv.** 1987. E6 and E7 gene products evolved by amplification of a 33-amino-acid peptide with a potential nucleic-acid-binding structure. Cancer Cells **5:**145–149.
7. **Delius, H., and B. Hofmann.** 1994. Primer-directed sequencing of human papillomavirus types. Curr. Top. Microbiol. Immunol. **186:**13–31.
8. **de Villiers, E. M.** 2001. Taxonomic classification of papillomaviruses. Papillomavirus Rep. **12:**57–63.
9. **Faccini, A. M., M. Cairney, G. H. Ashrafi, M. E. Finbow, M. S. Campo, and J. D. Pitts.** 1996. The bovine papillomavirus type 4 E8 protein binds to ductin and causes loss of gap junctional intercellular communication in primary fibroblasts. J. Virol. **70:**9041–9045.
10. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73:**237–244.
11. **Jackson, M. E., W. D. Pennie, R. E. McCaffery, K. T. Smith, G. J. Grindlay, and M. S. Campo.** 1991. The B subgroup bovine papillomaviruses lack an identifiable E6 open reading frame. Mol. Carcinog. **4:**382–387.
12. **Jacobson, E. R., C. R. Mladinich, S. Clubb, J. P. Sundberg, and W. D. Lancaster.** 1983. Papilloma-like virus infection in an African gray parrot. J. Am. Vet. Med. Assoc. **183:**1307–1308.
13. **Lina, P. H., M. J. van Noord, and F. G. de Groot.** 1973. Detection of virus in squamous papillomas of the wild bird species *Fringilla coelebs*. J. Natl. Cancer Inst. **50:**567–571.
14. **Mantovani, F., and L. Banks.** 2001. The human papillomavirus E6 protein and its contribution to malignant progression. Oncogene **20:**7874–7887.
15. **Moreno-Lopez, J., H. Ahola, A. Eriksson, P. Bergman, and U. Pettersson.** 1987. Reindeer papillomavirus transforming properties correlate with a highly conserved E5 region. J. Virol. **61:**3394–3400.
16. **Moreno-Lopez, J., H. Ahola, A. Stenlund, A. Osterhaus, and U. Pettersson.** 1984. Genome of an avian papillomavirus. J. Virol. **51:**872–875.
17. **Munger, K., J. R. Basile, S. Duensing, A. Eichten, S. L. Gonzalez, M. Grace, and V. L. Zacny.** 2001. Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. Oncogene **20:**7888–7898.
18. **O'Banion, M. K., E. R. Jacobson, and J. P. Sundberg.** 1992. Molecular cloning and partial characterization of a parrot papillomavirus. Intervirology **33:**91–96.
19. **O'Banion, M. K., M. E. Reichmann, and J. P. Sundberg.** 1986. Cloning and characterization of an equine cutaneous papillomavirus. Virology **152:**100–109.
20. **Osterhaus, A. D., D. J. Ellens, and M. C. Horzinek.** 1977. Identification and characterization of a papillomavirus from birds (*Fringillidae*). Intervirology **8:**351–359.
21. **Sundberg, J. P.** 1987. Papillomavirus infections in animals, p. 40–103. *In* K. Syrjanen, L. Gissmann, and L. G. Koss (ed.), Papillomaviruses and human disease. Springer-Verlag, Heidelberg, Germany.
22. **Sundberg, J. P., M. van Ranst, R. D. Burk, and A. B. Jenson.** 1996. The nonhuman (animal) papillomaviruses: host range, epitope conservation, and molecular diversity, p. 47–68. *In* G. Gross and G. von Krogh (ed.), Human papillomavirus infections in dermatology and venereology. CRC Press, Inc., Boca Raton, Fla.
23. **Van Bressem, M. F., K. Van Waerebeek, and J. A. Raga.** 1999. A review of virus infections of cetaceans and the potential impact of morbilliviruses, poxviruses and papillomaviruses on host population dynamics. Dis. Aquat. Org. **38:**53–65.
24. **Van Ranst, M., J. B. Kaplan, J. P. Sundberg, and R. D. Burk.** 1994. Molecular evolution of papillomaviruses. Cambridge University Press, Cambridge, United Kingdom.